

A Master equation approach to modeling an artificial protein motor

Nathan J. Kuwada^{a,*}, Gerhard A. Blab^{a,b}, Heiner Linke^{a,c}

^a*Department of Physics and Materials Science Institute, University of Oregon, 1274 University of Oregon, Eugene, OR, 97403-1274, USA*

^b*Department of Physics and IRMACS Centre, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada*

^c*The Nanometer Structure Consortium and Division of Solid State Physics, Lund University, Box 118, 22100 Lund, Sweden*

Abstract

Linear bio-molecular motors move unidirectionally along a track by coordinating several different processes, such as fuel (ATP) capture, hydrolysis, conformational changes, binding and unbinding from a track, and center-of-mass diffusion. A better understanding of the interdependencies between these processes, which take place over a wide range of different time scales, would help elucidate the general operational principles of molecular motors. Artificial molecular motors present a unique opportunity for such a study because motor structure and function are *a priori* known. Here we describe use of a Master equation approach, integrated with input from Langevin and molecular dynamics modeling, to stochastically model a molecular motor across many time scales. We apply this approach to a specific concept for an artificial protein motor, the Tumbleweed.

Keywords: Artificial Molecular Motors, Master equation, Computational Simulation

1. Introduction

Naturally occurring protein-based molecular motors are a broad class of macromolecules that transduce chemical potential energy into directed transport. Examples include rotary motors responsible for ATP synthesis (ATP-synthase) and locomotion (flagella), as well as linear motors, which move unidirectionally along a polymeric track and are responsible for, e.g., muscle contraction, cargo transport, and DNA/RNA transcription and replication [1, 2, 3]. Enabled by revolutionary advances in single molecule detection techniques, such as fluorescence microscopy and optical tweezers, the activity and performance of individual molecular motors has been directly observed (for an overview of techniques and accomplishments see [4]).

Inspired by biomolecular motors, several efforts to construct artificial molecular motors have been pursued for the last decade or so, one driving force being the vision of biomimetic, nanoscale machinery [5]. In addition, artificial molecular motors offer an opportunity to develop better understanding of the general operational principles of molecular motors, because of the

*Corresponding author, Tel 1-541-346-4583 Email address: nkuwada@uoregon.edu

design freedom offered by synthetic motors, and because motor structure and function typically are *a priori* known, allowing for detailed modeling.

One approach to artificial molecular motors is the use of synthetic, small molecules, which has led to a large variety of designs of motors and motor parts [6, 7, 8, 9] including a surface-mounted, photochemically driven, rotary motor where a combination of light-driven *cis-trans* isomerizations and thermal relaxations cause a molecule to make a complete 360° rotation in a predefined direction [10]. This particular design has also been subject of detailed modeling [11, 12], exemplifying this advantage of synthetic motors.

In a competing approach [13], and taking advantage of the highly designable, self-organized synthesis offered by oligonucleotides, Bath and Turberfield have constructed a linear DNA-based motor that is designed to step unidirectionally along a DNA-track [14]. The motor consists of two single-stranded (ss) DNA segments as ‘feet,’ flexibly joined by a double-stranded (ds) linker, that attach to complementary sequences on an ss-track. To achieve directional stepping, an ss-DNA fuel is introduced which selectively detaches the rear foot by competitive binding, allowing the rear foot to diffuse forward and complete a step.

In a very recent third approach, a design concept for an artificial, protein-based motor was proposed [15], with the aim to develop artificial motors that are based on the same material as biological motors. This design, the Tumbleweed, uses externally controlled, ligand-gated binding of repressor proteins to a DNA track to achieve unidirectional motion, and will be described in more detail below (Fig. 1).

Artificial and biomolecular motors alike must coordinate many processes to achieve unidirectional motion. In the case of linear DNA- and protein motors, these processes generally include fuel (ligand) capture, molecular conformational changes, binding to and unbinding from a track, and center-of-mass diffusion. The understanding of the interplay between these processes, which can span a vast range of time scales, is central to the design process of artificial motors, and to developing a better understanding of molecular motors in general.

Several different modeling approaches have been used to model molecular motors, each addressing different aspects of this many-time-scales problem. First, fundamental approaches to directed motion of small, Brownian molecules were developed based on spatially asymmetric ratchets [16, 17] and Brownian Motor models [18, 19], and helped understand the role of symmetry breaking, thermal noise and diffusion in molecular motors [20]. Such models are typically not designed to capture the molecule-specific *structure-function* relationship as well as details about sub-step processes of biological motors.

Molecular dynamics (MD) and Langevin dynamics models (LD), where molecular trajectories are calculated from an equation of motion, provide single-molecule information with potentially high accuracy, but require large amounts of computational time to simulate motor stepping, because all short-time processes must be explicitly calculated during the simulation run. One example of this approach are recent LD models of myosin V [21, 22, 23], where the modeling of motor stepping was enabled by coarse-graining of the molecule. Fully atomistic molecular dynamics (MD) simulations are currently computationally limited to simulation times on the order of nanoseconds, and thus too demanding for full simulations of motor stepping.

Finally, stochastic models, such as the Master equation approach described in more detail below, sacrifice the single-molecule information of dynamic models by modeling transitions between ensembles of motor states, but have the great advantage that they allow the modeling of motor processes across many different time scales [24].

In developing a complex artificial motor, one would thus like to combine the advantages of several of the above approaches, specifically the ability to relate atomistic detail to resulting

motor performance in terms of stepping rate and run length. Here we describe the integration of Master equation (ME) modeling with results from LD and MD simulations in the design process of the protein motor Tumbleweed [15]. In the following, we first describe the Tumbleweed concept in more detail, followed by a discussion of the time scales of a number processes that must be mutually tuned for stepping to result. Some of these time scales, such as the time for diffusional search for the next binding site, depend on motor design, are thus tunable to a degree, and can be estimated from LD modeling. Using a ME approach, enabled by our detailed knowledge of the possible motor states in this designed motor, we can explore the sensitivity of the motor design to chosen times and rates. The model output, such as quantitative relationships that must be fulfilled for successful motor performance, can then be used in the experimental motor design.

2. The Tumbleweed Motor

The Tumbleweed (TW) motor concept was developed as a first step towards artificial motors based on proteins [15]. It consists of three DNA-binding repressor proteins (R_A , R_B , R_C) attached to the corners of a Y-shaped, designed coiled-coil protein hub. The repressor proteins (the motor ‘feet’) have characteristically strong binding affinity to a unique ds-DNA base motif only in the presence of a specific ligand in solution (a , b , c). Thus the binding activity of each foot can be controlled separately by the external ligand supply. The track is then a designed length of ds-DNA with cyclic repeats of the three repressor protein binding motifs (binding sites). Stepping of the TW molecule is by center-of-mass diffusion, and directed transport is achieved by coordinated binding and unbinding of the motor feet, externally coordinated by cycling ligands in solution. The direction of transport is determined by both the binding site arrangement and the temporal order of the external ligand supply, and the speed of the motor is determined by how quickly ligands can be exchanged (see Fig. 1).

3. Modeling Tumbleweed: A Problem of Timescales

Consider TW in the state shown at the top of Fig. 1, with repressors R_A and R_B bound. To initiate a step, ligand a in solution is replaced by ligand c . For the step to take place, first the binding protein R_A must lose its ligand a , transitioning from its holo- (with ligand) to its apo-state (without ligand), and protein R_C must gain its ligand c . Apoprotein R_A must then unbind from the track, and the motor must diffuse until holoprotein R_C is near its next binding site and binds. Throughout this process, the external ligand supply must not be changed, and holoprotein R_B must stay attached to the track, or else the motor will fall off and be lost in the solution.

Qualitatively, these conditions are summarized by the inequality

$$\frac{1}{k_{\text{on/off}}} < \tau_{\text{diff}} < \tau_{\text{lig}} < \frac{1}{k_{\text{off}}^{\text{holo}}}, \quad (1)$$

in which $k_{\text{on/off}}$ are the ligand-repressor (dis)association rates, τ_{diff} is the diffusional stepping (searching) time of the motor, τ_{lig} is the time period during which the ligand concentration is kept constant, and $k_{\text{off}}^{\text{holo}}$ is the rate for a repressor holoprotein detachment from the DNA track.

A central modeling question for the design of TW molecule and experimental setup is: What are these different time scales and how will their absolute and relative values affect the performance of Tumbleweed?

The value of $1/k_{\text{on/off}}$ is not well known for our choice of repressor proteins, but is thought to be relatively fast compared to any other time scales in the system, on the order of picoseconds [25]. Modeling the system on this level requires a full atomistic molecular dynamics approach (MD), where the motion and interactions of atoms in the molecule as well as the surrounding ligands and buffer solution are explicitly calculated. The second term in Eq. 1 is the characteristic time for the TW molecule to rotationally diffuse to its next binding site. To get a rough idea of this term, the time scale for a sphere with the approximate diameter of the TW molecule (~ 20 nm) to diffuse one binding length (~ 10 nm) is around $1 \mu\text{s}$. To get a detailed value for this term, one can also use fully atomistic MD modeling, which allows one to consider steric hindrances between molecular components and to understand effective flexibilities of the hub-repressor protein joints. For current computational power, the full atomistic MD of the TW is able to provide maximum simulation runs of approximately 10^{-9} s, which is near the time scale for diffusion of subsections of the molecule, such as one arm of the central hub, but is not long enough to completely explore τ_{diff} . To fully explore how the stepping time depends on parameters such as joint flexibility or molecule size, we can coarse-grain the MD model to decrease computational time and increase the total simulation run-length time. Inertial motion of the TW molecule becomes completely damped by the surrounding fluid for times greater than 10^{-9} s, and thus can be neglected from the equations of motion for longer times. The atomistic picture of the molecule interacting with fluid molecules can then be replaced with a Langevin Equation, in which the fluid-molecule interactions are introduced as stochastic thermal noise and where molecular components are approximated by geometric objects whose sizes and viscous drag coefficients match the original motor components. This over-damped Langevin Dynamics (LD) approach can then use as input results from a MD model as described above, such as steric constraints and joint flexibilities, to correctly approximate the molecular components. For TW, the LD approach is applicable to time scales between μs and ms , which is right around the center-of-mass diffusion time of $\sim 150 \mu\text{s}$ [15].

The last two terms of Eq. 1, as well as the motor run lengths of interest, are in the 0.1-100 s range. Neither one of the dynamical models mentioned are able to efficiently explore this time scale. And because the dynamic models are tuned to model a specific range of time scales, they are not well equipped to explore the sensitivity of TW to the *interactions* of processes across different time scales. To quantitatively determine how sensitive TW is to the parameters in Eq. 1, which spans across many orders of magnitude, we must sacrifice single-molecule information and instead use a stochastic modeling technique that is able to model processes across many time scales: The Master equation.

4. The Master Equation Approach

The basis of the Master equation (ME) approach is the identification of motor states and their associated transition rates. In the case of the Tumbleweed, motor states are defined as distinct combinations of binding between the motor, its ligands, and the DNA track. For example, the motor state $A^*B\bar{C}^*$ represents the situation where foot R_A has its ligand bound but is not attached to the DNA, foot R_B has no ligand bound, and foot R_C has both its ligand bound and is attached to the DNA track ($*$ and $\bar{\cdot}$ thus represent ligand and DNA binding, respectively). In total, there are 80 distinct states for the TW motor: $2^3 = 8$ ligand binding states times 10 DNA-binding states, shown enumerated in Table 1 below. The number of possible DNA binding states is larger than the number of ligand binding states, as there are three theoretically possible configurations if all three feet are bound at the same time.

The ME is a differential equation that describes transitions between states. If all kinetics are approximated as first-order reactions, the ME is also of first order:

$$\frac{\partial}{\partial t} p(t) = M(t)p(t) \quad (2)$$

where (for a system with N-states) $p(t)$ is a N-dimensional vector with the numeric value of the n'th row representing the probability of finding the TW in the n'th motor state, and $M(t)$ is an NxN matrix of transition rates between states. A schematic representation of the allowed transitions for the TW motor is shown in Fig. 2. With the allowed transitions between the states defined, we can now form the $M(t)$ transition matrix. As shown in the state-naming convention of Table 1, it becomes convenient to define separate 8x8 sub-matrices L_i describing ligand exchange at constant DNA binding i , and D_j^i , which represent a change between DNA binding states i and j , respectively. The full $M(t)$ is then:

$$\begin{pmatrix} L_I & D_{II}^I & D_{III}^I & D_{IV}^I & 0 & 0 & 0 & 0 & 0 & 0 \\ D_{II}^I & L_{II} & 0 & 0 & D_V^II & D_{VI}^II & 0 & 0 & 0 & 0 \\ D_{III}^I & 0 & L_{III} & 0 & D_V^III & 0 & D_{VII}^III & 0 & 0 & 0 \\ D_{IV}^I & 0 & 0 & L_{IV} & 0 & D_{VI}^IV & D_{VII}^IV & 0 & 0 & 0 \\ 0 & D_V^V & D_{III}^V & 0 & L_V & 0 & 0 & D_{VIIa}^V & D_{VIIb}^V & D_{VIIc}^V \\ 0 & D_{II}^{VI} & 0 & D_{IV}^{VI} & 0 & L_{VI} & 0 & D_{VIIa}^{VI} & D_{VIIb}^{VI} & D_{VIIc}^{VI} \\ 0 & 0 & D_{III}^{VII} & D_{IV}^{VII} & 0 & 0 & L_{VII} & D_{VIIa}^{VII} & D_{VIIb}^{VII} & D_{VIIc}^{VII} \\ 0 & 0 & 0 & 0 & D_{V}^{VIIa} & D_{VI}^{VIIa} & D_{VII}^{VIIa} & L_{VIIa} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_{V}^{VIIb} & D_{VI}^{VIIb} & D_{VII}^{VIIb} & 0 & L_{VIIb} & 0 \\ 0 & 0 & 0 & 0 & D_{V}^{VIIc} & D_{VI}^{VIIc} & D_{VII}^{VIIc} & 0 & 0 & L_{VIIc} \end{pmatrix} \quad (3)$$

with the two types of sub-matrices defined as:

$$D_j^i = \begin{pmatrix} k_1^{ij} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_2^{ij} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_3^{ij} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_4^{ij} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_5^{ij} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_6^{ij} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_7^{ij} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_8^{ij} \end{pmatrix} \quad (4)$$

where k_ℓ^{ij} describes the transition between two DNA binding configurations i and j , while keeping the ligand binding configuration ℓ constant, and:

$$L_i = \begin{pmatrix} -\Sigma & k_{-A|A}^i & k_{-B|B}^i & k_{-C|C}^i & 0 & 0 & 0 & 0 \\ k_{+A|0}^i & -\Sigma & 0 & 0 & k_{-B|AB}^i & k_{-C|AC}^i & 0 & 0 \\ k_{+B|0}^i & 0 & -\Sigma & 0 & k_{-A|AB}^i & 0 & k_{-C|BC}^i & 0 \\ k_{+C|0}^i & 0 & 0 & -\Sigma & 0 & k_{-A|AC}^i & k_{-B|BC}^i & 0 \\ 0 & k_{+B|A}^i & k_{+A|B}^i & 0 & -\Sigma & 0 & 0 & k_{-C|ABC}^i \\ 0 & k_{+C|A}^i & 0 & k_{+A|C}^i & 0 & -\Sigma & 0 & k_{-B|ABC}^i \\ 0 & 0 & k_{+C|B}^i & k_{+B|C}^i & 0 & 0 & -\Sigma & k_{-A|ABC}^i \\ 0 & 0 & 0 & 0 & k_{+C|AB}^i & k_{+B|AC}^i & k_{+A|BC}^i & -\Sigma \end{pmatrix} \quad (5)$$

where $k_{-B|AB}^i$ is the associated rate for the process $A^*B^*C \Rightarrow A^*BC$ for the sustained DNA binding configuration i . The term $-\Sigma$ in the L_i matrix is the negative sum over the corresponding column of the full matrix, which is included to conserve the population. Each of the 10 L_i matrices contains 24 independent rates, and each of the 36 D_j^i matrices is fully defined by 8

independent rates, which results in a maximal set of 528 independent parameters to completely describe the dynamics of the TW motor. Despite the large number of parameters in the TW Master equation, the actual transition matrix $M(t)$ is actually rather sparsely populated and is best treated as such to allow for efficient and fast calculation. A strength of the ME approach is the inclusion of the wide range of relevant time scales in a single calculation. This strength, however, results in a stiff equation, i.e. the equation includes terms that can lead to rapid variation and numerical instability with most standard solvers, unless the time steps are chosen extremely small. We have chosen a specialized solver for stiff ordinary differential equations (`ode5r`, GNU Octave [26]) to allow us to solve the ME quickly and accurately.

5. Rates and States

Many of the compound rates needed to populate the transition matrix Eq. 5 are based on the component rates of free repressor proteins, but we must allow for the possibility that they are modified by the sterical constraints and effective forces generated by the connecting hub and the binding to the DNA recognition sites. Such modifications can at best be estimated by using LD or MD, but they are essentially not precisely known until a functioning prototype of the Tumbleweed motor has been constructed. For the moment we reduce the complexity of the problem by substituting those component rates known from literature for compound rates. In particular, we will consider the case where the ligand interaction L_i does not depend on the current DNA binding state of the repressor, in essence reducing the set of parameters to the quantities introduced in Eq. 1. The most easily experimentally accessible parameters are the DNA binding and release times $k_{\text{on}}^{\text{holo/apo}}$ and $k_{\text{off}}^{\text{holo/apo}}$ in the presence (holo-) and absence (apo-) of ligand, respectively. While detailed information in the literature is sparse, $k_{\text{off}}^{\text{holo}}$ is generally found on the order of $5 \cdot 10^{-3} \text{ s}^{-1}$ [27, 28, 29, 30], and is expected to be experimentally tunable by salt concentration. The corresponding $k_{\text{off}}^{\text{apo}}$ is at least 2-3 orders of magnitude greater, with preliminary experimental results placing it even higher, on the order of 10^3 s^{-1} [31, 32, 33]. The binding rate $k_{\text{on}}^{\text{holo}}$ depends on the relative concentration of repressor and DNA. In our simulations, we replace it by the diffusive search time τ_{diff} that we established by a Langevin simulation of the TW [15] and found to be about $\sim 150 \mu\text{s}$. For the remainder of this paper, we use the following values unless specified otherwise: $k_{\text{off}}^{\text{holo}} = 10^{-2} \text{ s}^{-1}$, $k_{\text{off}}^{\text{apo}} = 10^3 \text{ s}^{-1}$, $\tau_{\text{diff}} = 200 \mu\text{s}$, and $\tau_{\text{lig}} = 1 \text{ s}$.

		I ABC	II $\bar{A}BC$	III $A\bar{B}C$	IV $A\bar{B}\bar{C}$	V $\bar{A}BC$	VI $\bar{A}\bar{B}C$	VII $\bar{A}\bar{B}\bar{C}$	VIIIa $\bar{A}\bar{B}\bar{C}$	VIIIb $\bar{B}\bar{C}\bar{A}$	VIIIc $\bar{C}\bar{A}\bar{B}$
1	ABC	1	9	17	25	33	41	49	57	65	73
2	A^*BC										
3	AB^*C	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
4	ABC^*										
5	A^*B^*C										
6	A^*BC^*	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
7	AB^*C^*										
8	$A^*B^*C^*$	8	16	24	32	40	48	56	64	72	80

Table 1: Naming convention for the 80 distinct states of the Tumbleweed motor. Roman numerals indicate states with identical DNA-binding (10 types), while arabic numbers on the left indicate identical ligand binding (8 types). Alternatively the states can also be numbered consecutively from 1 to 80.

6. Results

A typical output from the Master equation model consists of the DNA binding probability for each foot as a function of time, as shown in Fig. 3. The ME described here does not yield spatial information, such as a trajectory or the speed of the TW, but instead only determines in which binding state a molecule would most likely be. Spatial information can be included in a ME simulation, but it requires one to define a unique set of states at each spatial coordinate, which will drastically increase the complexity of the problem and thus the computational time, depending on how long a track one wishes to include. But we can imply stepping information from our ME by examining specific state transitions throughout the ligand exchange cycle. For example, if the binding probability for R_A is zero during the entire $[b,c]$ ligand plug and then returns to one during the $[c,a]$ plug, the motors behavior is consistent with a state transition from state $\bar{A}\bar{B}\bar{C}$ to $\bar{A}B\bar{C}$, in agreement with the expected ligand controlled binding sequence $\bar{A}\bar{B}\bar{C} \rightarrow A\bar{B}\bar{C} \rightarrow \bar{A}B\bar{C} \rightarrow \bar{A}\bar{B}C$.

We are now equipped to return to a question raised by Eq. 1: For Tumbleweed to take successive steps and not fall off the track, what is the required quantitative relationship between the diffusional step time (τ_{diff}), the external ligand exchange time (τ_{lig}) and the average time for the repressor proteins to detach from the track ($1/k_{\text{off}}^{\text{holo}}$)?

τ_{lig} is determined by the time scale on which the ligand concentration can be changed microfluidically, and can be expected to be on the order of 0.1-1 s. Fig. 4 shows the population of motors attached to the track after 30 steps as a function of $k_{\text{off}}^{\text{holo}}$ for different values of τ_{diff} with τ_{lig} fixed at 1 s. For τ_{diff} on the 1- 100 μs time scale, the motor remains fairly successful even as $1/k_{\text{off}}^{\text{holo}}$ becomes less than τ_{lig} , but for $\tau_{\text{diff}} = 2 \text{ ms}$, we see a 20% reduction in motor attachment when $1/k_{\text{off}}^{\text{holo}} = \tau_{\text{lig}} = 1 \text{ s}$. It is initially surprising that the diffusive process can affect stepping success even though it is temporally separated from other relevant rates by 3 orders of magnitude. A physical interpretation of this finding is as follows: the primary reason for the motor to detach from the track is when $1/k_{\text{off}}^{\text{holo}}$ becomes less than τ_{lig} , because the motor is then simply not sticky enough. For short τ_{diff} however, the motor has a good chance to survive anyway, because it is attached to the track with two feet most of the time, providing a backup for one-foot detachment events. However, the vulnerability for detachment is amplified for increasing τ_{diff} , because the motor is attached with only one foot to the track for a greater fraction of the cycle.

We can also use the ME approach to investigate experimental design considerations and their consequences on the performance of TW. For example, there need to be at least three different buffer plugs containing ligands, $[a,b]$ $[b,c]$ $[c,a]$, microfluidically introduced in sequence into a microchamber containing the TW and immobilized DNA, to cycle the motor. As there is a finite length the plugs must travel before they reach the molecules, there will be some interdiffusion between plugs, creating pseudo-plugs containing all three ligands. Although, in the current molecular design, the TW molecule is sterically hindered from binding all three binding proteins at the same time (state $\bar{A}\bar{B}\bar{C}$ is not possible), the addition of triply mixed plugs reduces the temporal asymmetry of the system and could slow the motor down due to missteps. How sensitive is the TW to plug-mixing, and does the experimental design need to be adjusted to compensate? The expected experimental concentration profiles with interdiffusion between clean two-ligand plugs are shown qualitatively in Fig. 5(a). Although the ME is capable of modeling any ligand concentration profile, it is computationally expensive to model a continuous concentration function because the transition matrix $M(t)$ has to be recalculated at each program time step. Instead, we use the step-function approximated pseudo-plug profile shown in Fig. 5(b), which is characterized by τ_1 and τ_2 , the durations of the expected ‘clean’ plug and the pseudo-plug, respectively,

with $\tau_1 + \tau_2 = \tau_{\text{lig}}$.

Fig. 6 shows the probability for state $\bar{A}\bar{B}C$ during the $[b,c]$ ligand plug for multiple values of τ_2/τ_{lig} with $\tau_{\text{lig}} = 1$ s. Instead of the 0% binding probability for R_A throughout the entire $[b,c]$ plug characteristic for $\tau_2 = 0$ (Fig. 3), we now see a non-zero, increasing probability for state $\bar{A}\bar{B}C$ during the mixed pseudo-plug. Because the $\bar{A}\bar{B}C$ probability is 0% during the unmixed portion of the $[b,c]$ plug τ_1 , this profile implies two possible behaviors: (1) a fraction of motors completes the diffusional step forward to the $A\bar{B}\bar{C}$ binding state but then detaches R_C and returns to the $\bar{A}\bar{B}C$ state during τ_2 ; (2) motors release R_A but do not complete the step during τ_1 and return to state $\bar{A}\bar{B}C$. In either scenario, these motors fall out of phase with the ligand supply, resulting in the following binding sequence during the rest of the ligand cycle: $\bar{A}BC$ (original site) $\rightarrow \bar{A}\bar{B}C$ (misstep) $\rightarrow A\bar{B}\bar{C}$ (backstep to rearward site) $\rightarrow \bar{A}BC$ (original site). Therefore a single misstep causes the motor to temporally stall not just for a single ligand plug but for an entire ligand exchange cycle. With τ_{lig} held constant, the longer τ_2 , the less time the motors have to complete a step and the more opportunity they have to return to the previous binding state and stall. The maximum percentage of these misstepping motors is determined by the fraction of motors in the $\bar{A}\bar{B}C$ state at the ligand exchange time ($t = 2$ s in Fig. 6) and is shown in the inset of Fig. 6.

The maximum stepping rate of TW is determined by τ_{lig} . What is the effect of ligand mixing if we chose to use a shorter τ_{lig} to increase the stepping rate? The inset of Fig. 7 shows the probability for state $\bar{A}\bar{B}C$ during the plugs $[a,b]$ and $[b,c]$ as a function of τ_2/τ_{lig} with $\tau_{\text{lig}} = 0.1$ s, an order of magnitude less than Fig. 6. Instead of the probability for state $\bar{A}BC$ going to 0% during τ_1 as we saw in Figs. 3 and 6 for $\tau_{\text{lig}} = 1$ s, we now see that for $\tau_2/\tau_{\text{lig}} > 0.5$ a fraction of motors remain in state $\bar{A}\bar{B}C$ throughout the entire pulse. This effect occurs when τ_1 becomes similar to the characteristic fall-off time of the $\bar{A}\bar{B}C$ state ($\sim 1/k_{\text{off}}^{\text{apo}}$), so a fraction of motors have not yet released R_A before the mixed pseudo-plug arrives and thus remain in state $\bar{A}\bar{B}C$. Because of the high misstepping percentage for $\tau_2/\tau_{\text{lig}} > 0.5$, we also begin to see a reduction in the probability for the expected state $\bar{A}BC$ during the $[a,b]$ ligand plug. The misstepping fraction for $\tau_{\text{lig}} = 0.1$ s is shown in Fig. 7. Therefore, although decreasing τ_{lig} has the possibility of increasing the speed of TW, the system also becomes more sensitive to high levels of mixing that could potentially *decrease* the motor speed by increasing misstepping events. Detailed ME modeling of these interdependencies will be highly useful for motor optimization.

7. Conclusions

We have shown how a Master equation approach can be integrated with input from molecular dynamics and coarse-grained Brownian dynamics (Langevin) modeling to efficiently explore the dependence of the performance of a synthetic molecular motor on variables on a large range of time scales.

For the specific case of the protein-motor concept Tumbleweed, returning to Eq. 1, we have seen that the successive stepping success of TW is not remarkably sensitive to τ_{diff} unless $1/k_{\text{off}}^{\text{holo}}$ and τ_{lig} are nearly equal. In this case, for an experimentally realistic $\tau_{\text{lig}} = 1$ s, the value of τ_{diff} can decrease motor attachment by 20% even if it is three orders of magnitude less than τ_{lig} . Although the component holoprotein detachment rates ($k_{\text{off}}^{\text{holo}}$) are expected to be $\sim 5 \cdot 10^{-3}$ s⁻¹ (suggesting that $1/k_{\text{off}}^{\text{holo}} \gg \tau_{\text{lig}}$), the compound rates may vary due to steric constraints, and to be safe the design of the TW molecule should be tuned to decrease τ_{diff} as much as possible. Future MD modeling will be used to determine how design choices affect τ_{diff} , e.g. through joint flexibilities and steric hindrances.

We have also determined how ligand plug mixing affects the stepping behavior of TW. The binding state behavior suggests that plug mixing can lead to missteps and temporary stalling events if the motor returns to its previous binding state during the mixed plug time τ_2 and consequently falls out of phase with the periodic ligand supply. Reducing τ_{lig} may, in principle, increase the speed of TW, but we have found the system also becomes more sensitive to ligand mixing, and for high levels of mixing temporary stalling events may actually decrease the speed, presenting an optimization challenge. For $\tau_{\text{lig}} = 1$ s, a mixing time $\tau_2/\tau_{\text{lig}} = 0.5$ gives a misstepping probability of $\sim 2.5\%$ per step. Current experimental designs for the TW motor have the possibility of reducing τ_2/τ_{lig} to ~ 0.05 where the probability of misstepping is negligible, thus the ligand mixing should not present a significant problem to the successful function of the Tumbleweed motor.

The authors would like to thank the HFSP Motor Collaboration for inspiration and useful discussions, specifically Martin J. Zuckermann for technical advice. This work is supported by the National Science Foundation under Grant Nos. DGE-0742540 and DGE-0549503 (NJK), and the Human Frontier Science Program (RGP0031/2007).

- [1] J. Howard. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer, Massachusetts, 2001.
- [2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 5th edition, 2008.
- [3] G. Banting and S. Higgins, editors. *Essays in Biochemistry: Molecular Motors*. Portland Press, London, 2000.
- [4] W. J. Greenleaf, M. T. Woodside, and S. M. Block. High-resolution, single-molecule measurements of biomolecular motion. *Annual Review of Biophysics and Biomolecular Structure*, 36:171–190, 2007.
- [5] B. L. Feringa. The art of building small: From molecular switches to molecular motors. *Journal of Organic Chemistry*, 72(18):6635–6652, 2007.
- [6] G. S. Kottas, L. I. Clarke, D. Horinek, and J. Michl. Artificial molecular rotors. *Chemical Reviews*, 105(4):1281–1376, 2005.
- [7] V. Balzani, A. Credi, S. Silvi, and M. Venturi. Artificial nanomachines based on interlocked molecular species: recent advances. *Chemical Society Reviews*, 35(11):1135–1149, 2006.
- [8] E. R. Kay, D. A. Leigh, and F. Zerbetto. Synthetic molecular motors and mechanical machines. *Angewandte Chemie-International Edition*, 46(1-2):72–191, 2007.
- [9] S. Silvi, M. Venturi, and A. Credi. Artificial molecular shuttles: from concepts to devices. *Journal of Materials Chemistry*, 19(16):2279–2294, 2008.
- [10] R. van Delden, M. ter Wiel, M. Pollard, J. Vicario, N. Koumura, and B. L. Feringa. Unidirectional molecular motor on a gold surface. *Nature*, 437(7063):1337–1340, 2005.
- [11] M. Klok, W. R. Browne, and B. L. Feringa. Kinetic analysis of the rotation rate of light-driven unidirectional molecular motors. *Physical Chemistry Chemical Physics*, 11(40):9124–9131, 2009.
- [12] E. M. Geertsema, S. J. van der Molen, M. Martens, and B. L. Feringa. Optimizing rotary processes in synthetic molecular motors. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40):16919–16924, 2009.
- [13] B. Yurke, A. J. Turberfield, A. P. Mills, F. C. Simmel, and J. L. Neumann. A dna-fuelled molecular machine made of dna. *Nature*, 406(6796):605–608, 2000.
- [14] J. Bath, S. J. Green, K. E. Allen, and A. J. Turberfield. Mechanism for a directional, processive, and reversible dna motor. *Small*, 5(13):1513–1516, 2009.
- [15] E. H. C. Bromley, N. J. Kuwada, M. J. Zuckermann, R. Donadini, L. Samii, G. A. Blab, G. J. Gemmen, B. J. Lopez, P. M. G. Curmi, N. R. Forde, D. N. Woolfson, and H. Linke. The tumbleweed: towards a synthetic protein motor. *HFSP Journal*, 3(3):204–212, 2009.
- [16] M. O. Magnasco. Forced thermal ratchets. *Physical Review Letters*, 71(10):1477–1481, 1993.
- [17] R. D. Astumian and M. Bier. Fluctuation driven ratchets - molecular motors. *Physical Review Letters*, 72(11):1766–1769, 1994.
- [18] R. D. Astumian. Thermodynamics and kinetics of a brownian motor. *Science*, 276(5314):917–922, 1997.
- [19] R. D. Astumian and P. Hanggi. Brownian motors. *Physics Today*, 55(11):33–39, 2002.
- [20] R. D. Astumian. Design principles for brownian molecular machines: how to swim in molasses and walk in a hurricane. *Physical Chemistry Chemical Physics*, 9(37):5067–5083, 2007.
- [21] E. M. Craig and H. Linke. Mechanochemical model for myosin v. *Proceedings of the National Academy of Sciences of the United States of America*, 106(43):18261–18266, 2009.
- [22] A. Vilfan. Five models for myosin v. *Frontiers in Bioscience*, 14:2269–2284, 2009.

- [23] A. Vilfan. Elastic lever-arm model for myosin v. *Biophysical Journal*, 88(6):3792–3805, 2005.
- [24] C. Bustamante, D. Keller, and G. Oster. The physics of molecular motors. *Accounts of Chemical Research*, 34(6):412–420, 2001.
- [25] O. Schaad, H. X. Zhou, A. Szabo, W. A. Eaton, and E. R. Henry. Simulation of the kinetics of ligand-binding to a protein by molecular dynamics - geminate rebinding of nitric-oxide to myoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, 90(20):9547–9551, 1993.
- [26] J. W. Eaton. *GNU Octave Manual*. Network Theory Limited, 2002.
- [27] I. D. Parsons, B. Persson, A. Mekhalfia, G. M. Blackburn, and P. G. Stockley. Probing the molecular mechanism of action of co-repressor in the e.coli methionine repressor-operator complex using surface plasmon resonance (SPR). *Nucleic Acids Research*, 23(2):211–216, 1995.
- [28] P. G. Stockley, A. J. Baron, C. M. Wild, I. D. Parsons, C. M. Miller, C. A. Holtham, and S. Baumberg. Dissecting the molecular details of prokaryotic transcriptional control by surface plasmon resonance: the methionine and arginine repressor proteins. *Biosensors and Bioelectronics*, 13(6):637–650, Sept. 1998.
- [29] M. D. Finucane and O. Jardetzky. Surface plasmon resonance studies of wild-type and AV77 tryptophan repressor resolve ambiguities in super-repressor activity. *Protein Science*, 12(8):1613–1620, 2003.
- [30] Y. Wang, L. Guo, I. Golding, E. C. Cox, and N. Ong. Quantitative transcription factor binding kinetics at the Single-Molecule level. *Biophysical Journal*, 96(2):609–620, 2009.
- [31] J. Carey. Gel retardation at low ph resolves trp repressor-dna complexes for quantitative study. *Proceedings of the National Academy of Sciences of the United States of America.*, 85(4), 1988.
- [32] K. Borden, P. Beckmann, and A. Lane. Determination of the orientations of tryptophan analogues bound to the trp repressor and the relationship to activation. *European Journal of Biochemistry*, 202(2), 1991.
- [33] K. Phillips and S. V. Phillips. Electrostatic activation of escherichia-coli methionine repressor. *Structure (London)*, 2(4), 1994.

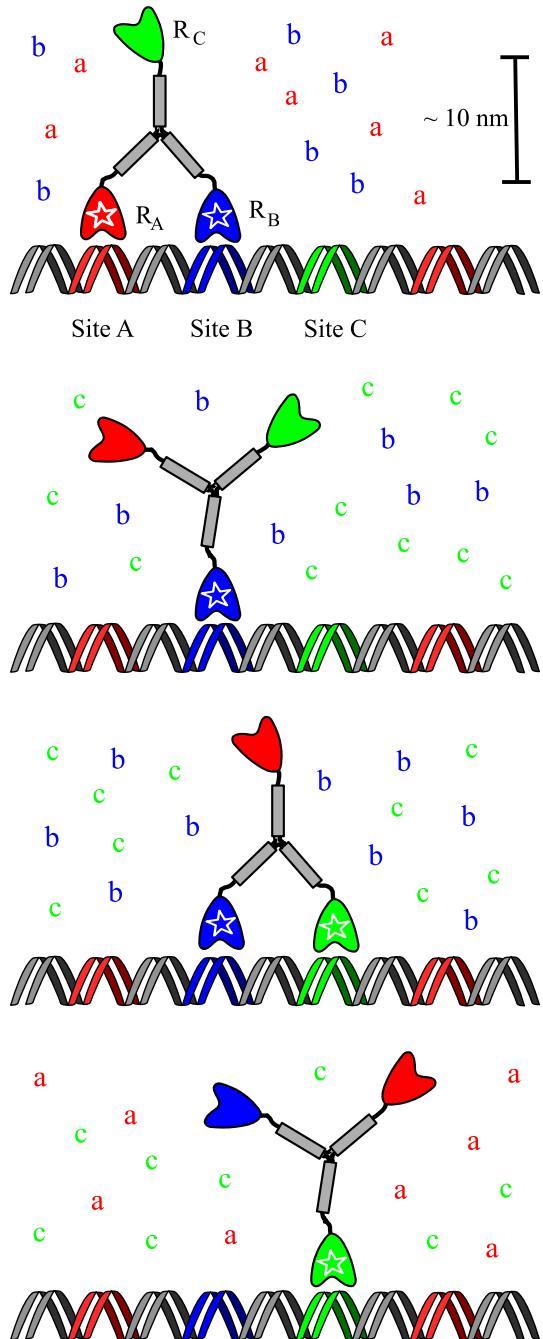


Figure 1: Stepping process of the Tumbleweed. A coordinated, directed stepping of the TW motor is achieved by control of the ligand supply (letters a, b, and c), and DNA binding site arrangement. Each repressor protein foot (R_A , R_B , and R_C , shown in red, blue and green, respectively) is only able to bind strongly to its corresponding binding site while it has its specific ligand bound (indicated by a white star). Ligands are offered in a temporally repeating sequence ("plugs") $[a,b]$, $[b,c]$, $[c,a]$. For the motor to successfully complete a step, the time scales for ligand capture, motor diffusion, and DNA binding must obey Eq. 1.

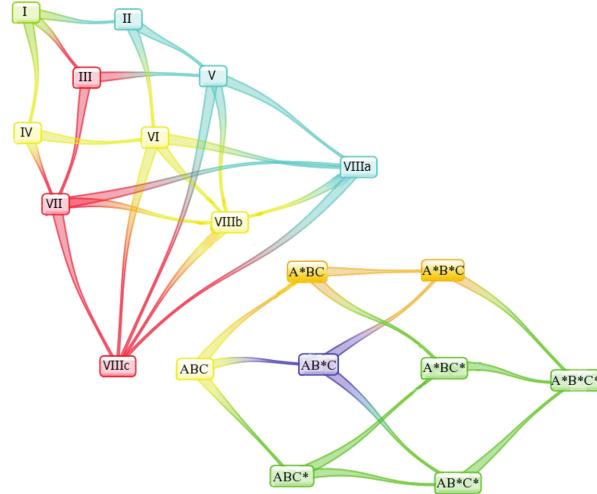


Figure 2: Visual representation of the allowed transitions of the Tumbleweed motor. An allowed transition can either change the DNA binding state (indicated by roman numerals, top) or the ligand binding configuration (bottom, a star indicates the respective repressor has its ligand bound). A complete list of states is given in Table 1

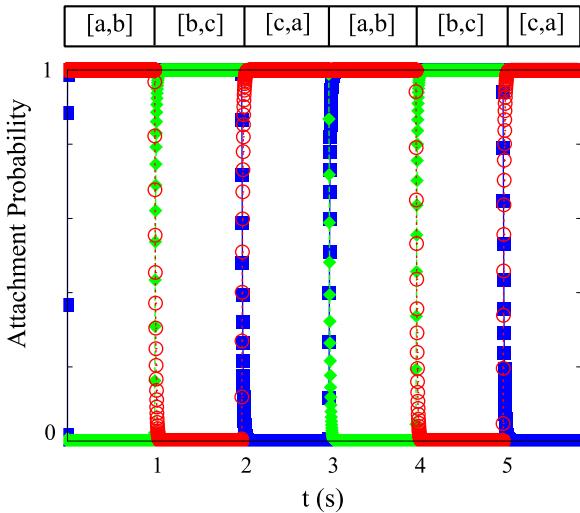


Figure 3: Attachment of the TW feet (R_A : red circle, R_B : blue square, R_C : green diamond) to their specific DNA site over two complete ligand cycles. The ligand plugs are indicated over the graph. The values for $\tau_{\text{diff}} = 200 \mu\text{s}$ and the detachment rate $1/k_{\text{off}}^{\text{holo}} = 100 \text{ s}$ are chosen such that the motor is expected to complete successive steps. We can infer stepping motion by the binding behavior: during the entire $[b,c]$ ligand plug (for $t = 1 \text{ s} - 2 \text{ s}$), we see 100% binding of R_B and R_C , while R_A is 0%, but R_A returns to 100% binding probability during the following $[c,a]$ plug ($t = 2 \text{ s} - 3 \text{ s}$) while R_B is 0%, which is consistent with the ligand controlled stepping behavior.

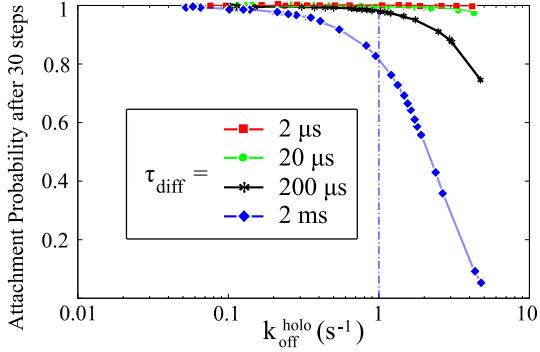


Figure 4: Successive stepping, measured by the percentage of motor still attached to the DNA after 30 steps, for four values of diffusive search time (τ_{diff}) as a function of detachment rate $k_{\text{off}}^{\text{holo}}$. We find that the diffusive search time must be several orders of magnitude shorter than the average time for detachment for the motor to remain bound to DNA. For all curves the ligand plug length is constant ($\tau_{\text{lig}} = 1 \text{ s}$).

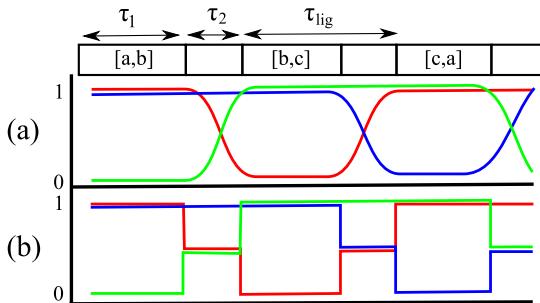


Figure 5: Interdiffusion between clean, binary ligand plugs of duration (τ_{lig}) results in the profile shown in (a), with time periods during which all three ligands (a , b , and c) are present in solution at the same time. For our numerical computations we use a simplified plug profile with discrete levels (b), characterized by a clean, binary plug of duration τ_1 ($< \tau_{\text{lig}}$), and a mixed ‘pseudo-plug’ of length τ_2 . The ratio τ_2/τ_{lig} defines the degree of the mixing.

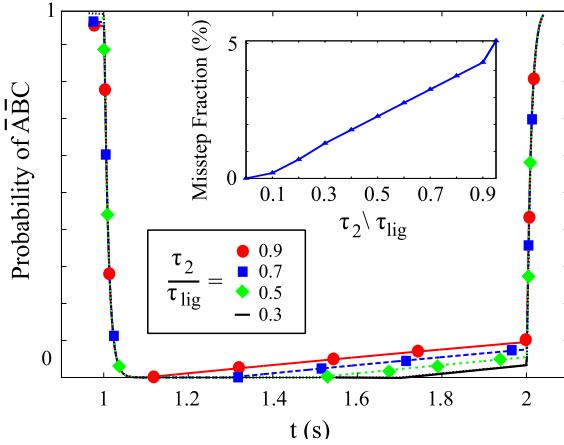


Figure 6: Attachment probability for state $\bar{A}\bar{B}C$ during the $[b,c]$ ligand plug for four values of τ_2/τ_{lig} with $\tau_{\text{lig}} = 1$ s. The binding probability profile here implies that a fraction of the motors are either completing the diffusional step to the $\bar{A}\bar{B}C$ state but then detaching R_C and returning to the previous $\bar{A}BC$ state, or simply not completing the step. Inset: The percentage of motors that misstep when ligands nominally are exchanged (time $t = 2$ s) as a function of τ_2/τ_{lig} .

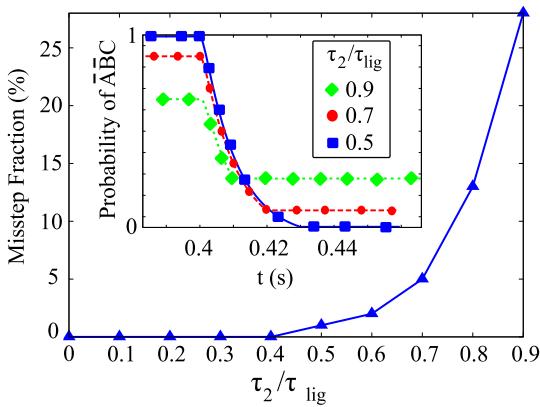


Figure 7: Fraction of misstepping motors after ligand plug exchange $[a,b]$ to $[b,c]$ as a function of τ_2/τ_{lig} with $\tau_{\text{lig}} = 0.1$ s. Inset: Probability of state $\bar{A}\bar{B}C$ for the $[a,b]$ and $[b,c]$ ligand plugs as a function of time for three values of τ_2/τ_{lig} . As the mixing time τ_2 is increased, the probability of the state $\bar{A}\bar{B}C$ during the $[b,c]$ ligand plug no longer approaches zero and increases with higher τ_2/τ_{lig} . Misstepping occurs when $\tau_2/\tau_{\text{lg}} > 0.5$, where τ_2 becomes comparable to $1/k_{\text{off}}^{\text{apo}}$ (0.001 s) and a fraction of the motors never releases R_A before the mixed pseudo-plug arrives. The high misstepping percentage also reduces the expected 100% probability for state $\bar{A}BC$ during the $[a,b]$ ligand plug.